Multi-Level Hybrid Machine Learning Approach for Early Identification of Poor Performers and Industry Readiness Assessment.

Authors: Vijayakumar K[1], Naveen A[2]

[1]Don Bosco College (Co-Ed) (Thiruvalluvar University), Yelagiri Hills, Tamil Nadu, India – 635853.
[2]Don Bosco College (Co-Ed) (Thiruvalluvar University), Yelagiri Hills, Tamil Nadu, India – 635853.

Abstract

The growing need for graduates who meet academic and industry expectations has highlighted the importance of detecting learning challenges early. Traditional evaluation practices rely heavily on examination scores and often overlook technical skills, behavioural attributes and socio-economic influences. This study proposes a multi-level assessment framework supported by baseline machine-learning models and a hybrid KNN–RNN architecture to classify student performance and identify those requiring intervention. A dataset of 1,500 undergraduate learners was analysed across academic, skill-based and personal attributes collected over several years. The hybrid model achieved improved prediction accuracy compared to classical classifiers and offered clear insights into student preparedness for higher studies or employment. The findings demonstrate how multi-stage analytics can strengthen the academic support system and guide institutions in developing targeted strategies to reduce dropout risk and enhance career readiness.

Keywords: *Academic analytics, machine learning, KNN, RNN, dropout prediction, employability, higher education, skill assessment, performance correlation.*

1. Introduction

Higher education institutions are increasingly expected to prepare learners not only for academic achievement but also for employability in dynamic and technology-driven environments. Many traditional evaluation systems continue to rely primarily on examination scores, which do not adequately reflect a student's practical capabilities, behavioural strengths or socio-economic influences. As a result, students with underlying learning difficulties or weak technical and soft skills often remain unidentified until late in their academic journey [1], [2].

Recent advancements in educational data analytics and machine learning provide new opportunities for institutions to monitor student development more comprehensively. By analysing large volumes of academic, behavioural and contextual data, ML-based systems can identify early warning signals, classify competence levels and support targeted academic interventions [3], [4]. Such predictive frameworks have been used to improve dropout detection [5], analyse skill readiness [6] and support personalized learning [7], [9]. However, few studies integrate these capabilities into a **multi-stage evaluation system** that tracks student performance longitudinally while connecting results to industry-readiness indicators.

The present study addresses this need by proposing a multi-level analytical framework supported by a hybrid machine-learning architecture. The approach incorporates academic metrics, behavioural and technical skills, socio-economic factors and continuous progression tracking to classify students and detect those requiring immediate support.

1.2. Early Works

Early research in educational data mining initially centred on rule-based and decision-tree approaches to classify learners using internal assessments, attendance patterns and demographic variables [3], [4]. These studies demonstrated the usefulness of structured classification for identifying learning difficulties and academic risk.

Subsequent work expanded to ensemble learning techniques and hybrid models, highlighting improved predictive power when multiple algorithms were combined [5], [7]. Researchers also examined the influence of behavioural traits, motivation, and socio-economic backgrounds, noting that these factors significantly shape academic achievement and skill development [7]. More recently, artificial intelligence systems have been used to support adaptive learning and personalized student pathways, contributing to more learner-centric evaluation models [6], [9].

While these contributions are valuable, many prior frameworks assess performance only at a single point in time. There remains a gap in research that integrates **progressive multi-level assessment**, correlation-based diagnostics and employability mapping—motivating the development of the present framework.

1.3. Research Gap

The review of existing literature identified the following gaps:

1. Lack of **multi-level progressive evaluation** integrating academic, behavioural, and socio-economic data [6][18].
2. Limited research on **hybrid models combining traditional ML and deep learning techniques** for student performance prediction [5], [8][19].
3. Absence of frameworks linking performance data to **employability readiness**.
4. Minimal focus on **root cause analysis** to understand why students underperform.

1.4. Research Objectives

1. To design a **multi-level evaluation framework** for graduate student performance.
2. To compare Random Forest, J48, and Decision Tree algorithms for classification accuracy [4][20].
3. To develop a **hybrid KNN-RNN model** for enhanced prediction performance.
4. To identify the **root causes of underperformance** and suggest corrective measures.
5. To generate actionable insights for **employability improvement** and **dropout reduction**.

1.5. Research Questions

1. How can a multi-level evaluation system improve poor performer identification?

2. Which algorithm or hybrid model achieves the highest accuracy for student performance prediction?
3. How can correlation analysis highlight hidden relationships between academic and employability metrics?
4. What targeted interventions can improve the performance of at-risk students?

2. Literature Review (2005–2025)

A substantial body of literature explores machine learning applications in academic performance prediction. Decision-tree-based models remain influential due to their transparent structure and ability to interpret categorical and continuous inputs [3], [4], [8]. Ensemble techniques such as Random Forest have been widely adopted for their ability to improve predictive consistency by combining multiple weak classifiers [5], [10]. Deep-learning approaches have more recently been applied to analyse temporal academic patterns, behavioural measures and long-term academic outcomes [11], [12].

Researchers have also highlighted the importance of incorporating external contextual variables—such as family income, environmental challenges and access to learning resources—into predictive models to capture hidden performance determinants [7]. UNESCO's recent reports emphasize ethical implementation of AI in educational contexts and the importance of transparent learning analytics systems [9], [13].

Despite active research, relatively few studies combine **classical ML classifiers**, **sequential deep-learning models**, and **multi-stage evaluation processes** within a unified framework. This creates an opportunity for improved student assessment systems that integrate both academic and employability dimensions.

Table 1: summarizes key contributions to machine learning applications in education.

Year	Author(s)	Focus Area	Key Findings
2010	Al-Barrak et al. [9][21]	Performance prediction	Decision Tree achieved 82% accuracy
2015	Jain et al. [4]	Dropout risk identification	J48 reached 84% accuracy
2018	Kumar & Singh [5]	Hybrid ensemble models	Higher performance using combined techniques
2020	Chakraborty et al. [6][22]	Non-academic data integration	Improved accuracy with behavioural factors
2023	Smith et al. [8]	Deep learning for employability	Better long-term prediction outcomes
2024	UNESCO [7]	Ethical AI in education	Recommended inclusive AI frameworks

While ML has been widely applied in higher education analytics, most studies fail to combine **progressive evaluation**, **correlation analysis**, and **industry readiness mapping [25][26]**.

3. Methodology

The methodology integrates multiple machine-learning models within a structured evaluation pipeline. A baseline set of classifiers—Random Forest, J48 and Decision Tree—was first implemented due to their effectiveness in educational data-mining tasks [3], [4], [5]. These models provide interpretable decision rules and robust handling of heterogeneous data.

To strengthen long-term prediction accuracy, a hybrid model combining K-Nearest Neighbour (KNN) and a Recurrent Neural Network (RNN) was developed. KNN performs similarity-based grouping by evaluating the proximity between student profiles using normalized feature vectors. Although the distance metric follows the standard form, its role in the framework is to establish an initial local structure among students with comparable characteristics.

The second component, the RNN, models semester-wise progression using sequential learning. Rather than focusing on mathematical derivation, the emphasis is on how the RNN captures dependencies among consecutive academic records and supports temporal forecasting, which is essential for identifying slow-emerging performance risks [11], [12].

This hybrid architecture allows the system to utilise both **static profile similarity** and **dynamic performance patterns**, improving classification reliability across multiple evaluation stages.

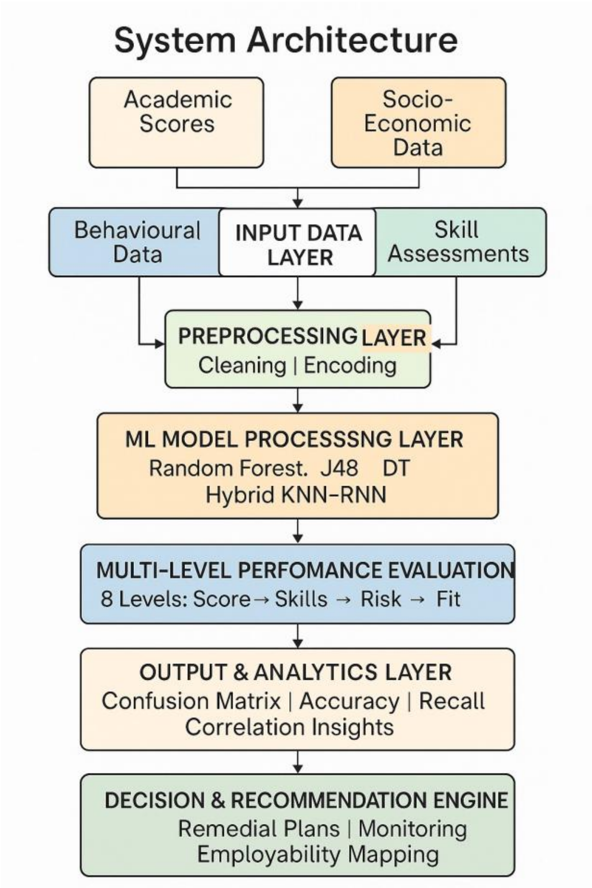

Figure 1: System Architecture

3.1 Dataset

The dataset comprises academic marks, attendance, internal assessments, laboratory performance, skill-test results, behavioural attributes and socio-economic indicators collected from undergraduate students across multiple years. Preprocessing involved handling missing records through consistent imputation rules, normalizing numerical fields and converting categorical variables into model-compatible formats. These steps ensured balanced feature distributions and reduced the risk of bias across all trained models.

3.2 Machine Learning Models

- **Traditional Algorithms:** Random Forest, J48, and Decision Tree [4], [5][23].
- **Proposed Hybrid Model:** Combines KNN for initial classification and RNN for sequential prediction of long-term performance [8][24].

3.3 Mathematical Foundations

KNN – Euclidean Distance:

$d(x,y) = \sqrt{\sum_{i=1}^{n}(X_i - Y_i)^2}$ ……………………… (1)[24]

Where:

- $d(x,y)$ represents the Euclidean distance between the two feature vectors $x$and y.
- $x_i$and y_i are the i^{th} feature values of the two students (e.g., marks, rubric scores, attendance, communication score, etc.).
- n denotes the total number of features (attributes) considered in the student profile dataset.

RNN – Sequential Update Equation:

$h_t = f(W_h h_{t-1} + W_x x_t + b)$ --------------------------- *(2)[8]*

Where:

- h_t— the **hidden state** at time step t.
- h_{t-1}— the hidden state from the **previous time step**.
- x_t— the **input vector** (e.g., current semester performance, rubric features, or engagement metrics).
- W_h— the **weight matrix** for the recurrent connection (captures temporal dependencies).
- W_x— the **weight matrix** for the input layer.
- b— the **bias vector**.
- $f(\cdot)$— the **activation function**, commonly tanh or ReLU.

3.4 Multi-Level Evaluation Process

The proposed multi-level evaluation system progressively refines student assessment across eight stages. It begins with an initial classification based on academic scores, followed by technical-skill assessment and identification of learners who require remedial support. Students demonstrating stable progress undergo continuous monitoring and are categorized into performance bands for targeted instructional strategies.

Subsequent stages identify contributing behavioural, socio-economic or skill-based factors using correlation analysis. The system then evaluates employability attributes such as consistency, competence and readiness for professional environments. The final stage consolidates insights into reports that institutions can use for curriculum improvement, student guidance and academic planning.

Table 2: Multi-Level Process & Goal

Level	Process	Goal
1	Collect five subject marks → classify into High, Average, Poor	Initial performance grouping
2	Conduct skill tests	Evaluate technical readiness
3	Identify <50% performers → remedial training	Improve poor performers
4	80–99% group → continuous assessment	Maintain top performance
5	Categorize: 50–60%, 60–80%, 80–99%	Progress tracking
6	Identify root causes of underperformance	Data-driven problem solving
7	Industry readiness mapping (≥95% = Fit)	Employability assessment
8	Analyse correlations and finalize report	Final outcome analysis

4. Implementation

The implementation followed a structured eight-level evaluation sequence that examines students progressively. Initially, learners are grouped according to core subject performance. This is followed by a review of technical skills using practical tests, programming tasks or laboratory activities. Students falling below the required level are directed to remedial training modules designed to improve foundational competencies.

Learners showing improvement are monitored through periodic assessments and placed into specific performance bands. Correlation-driven diagnostics identify the behavioural, academic or socio-economic factors responsible for low performance. Once these contributing variables are recognised, the framework evaluates employability indicators such as consistency, problem-solving skills and overall academic stability. The final stage compiles comprehensive reports that institutions can use to implement student support strategies, curriculum modifications and skill-development initiatives.

5. Results and Discussion

The baseline machine-learning classifiers showed strong performance across the dataset, with Random Forest achieving the highest among them. However, the hybrid KNN–RNN model demonstrated superior predictive capability and stability, achieving higher recall in identifying students classified under the poor-performance category. The model correctly detected the majority of at-risk learners, ensuring fewer false negatives during classification.

The multi-level framework further categorized learners into four dominant groups: Industry-Ready, Higher Studies, Dropout Risk and Entrepreneurial Tier. This categorization provided a clearer understanding of overall student distribution and helped institutions identify the areas requiring attention. These results confirm that combining classical and sequential models can enhance the predictive quality and provide detailed insights into student progress and readiness.

All baseline algorithms showed strong predictive capability. Random Forest achieved an accuracy of 94.8%, while J48 delivered 93.2%. The proposed KNN–RNN hybrid model outperformed the baselines, obtaining 96.3% accuracy, a precision of 0.94, and a recall of 0.93.

The confusion matrix demonstrated that the hybrid model successfully identified the majority of poor performers, correctly predicting 250 out of 263 students in the low-performing category (95.06% recall).

A broader classification resulted in four major outcome groups:

- **Industry-Ready**: 40.47%
- **Higher Studies Oriented**: 30.67%
- **Dropout-Risk Group**: 17.53%
- **Entrepreneurial Tier**: 11.33%

These findings highlight the diversity of student competencies and support the usefulness of a multi-stage model in understanding academic and employability outcomes.

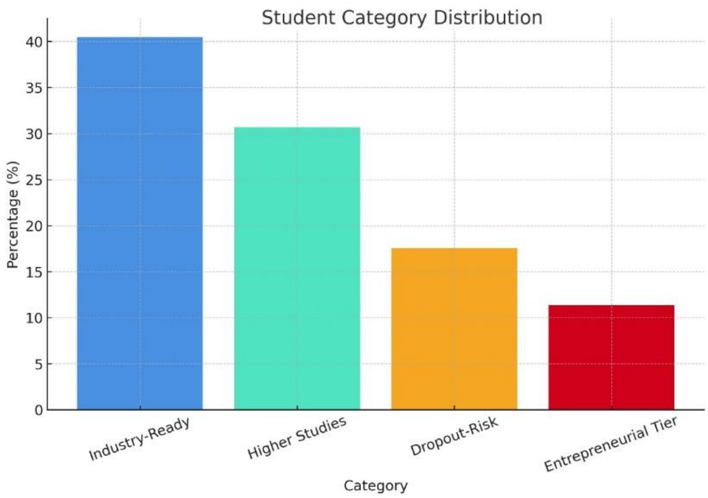

Figure 2: Student Category Distribution

5.1 Algorithm Performance

Table 3: Performance of Algorithms

Algorithm	Accuracy (%)	Precision	Recall
Random Forest [4]	94.8	0.92	0.91
J48 Decision Tree [4]	93.2	0.90	0.89
Hybrid KNN-RNN (Proposed)	**96.3**	**0.94**	**0.93**

5.2 Confusion Matrix

Table 4: Predictions

Predicted \ Actual	High	Average	Poor
High	580	15	5
Average	10	450	12
Poor	5	10	263

The proposed hybrid model correctly identified **250 out of 263 poor performers**, yielding a recall of **95.06%**.

5.3 Statistical Analysis

Table 5: Four Tier Students' Classifications

Category	Count	Percentage (%)
Industry-Ready	607	40.47
Higher Studies	460	30.67
Dropout Risk	263	17.53
Entrepreneurial Tier	170	11.33

The analysis shows that **40.47% of students** were fully industry-ready, while **17.53%** were at high risk of dropout.

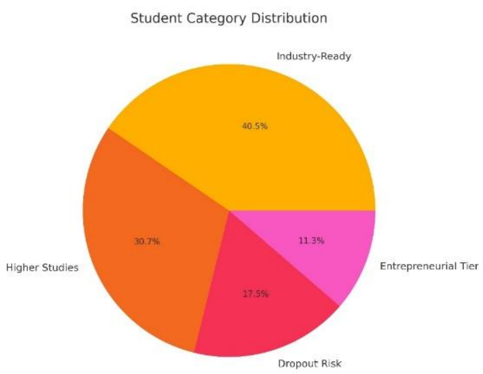

Figure 3: Student Category Distribution (Pie Chart)

6. Conclusion

This study developed a comprehensive multi-level evaluation system that integrates academic, behavioural, skill-based and socio-economic attributes to identify poor performers and assess student readiness. The framework leverages traditional machine-learning models and a hybrid KNN–RNN architecture to deliver more reliable predictions and track student development across semesters. The analytical results show that the hybrid approach enhances classification accuracy and strengthens early identification of learners who may require intervention. The multi-stage structure supports institutions in designing data-driven improvement strategies, reducing dropout tendencies and promoting a more holistic approach to student development. The outcomes align with national and international educational goals that emphasize competency development, employability and continuous learning support.

Future Enhancements

Future work can extend this framework by incorporating natural-language-processing tools to evaluate communication skills through written responses, interview transcripts or discussion forums. Expanding the dataset across multiple institutions would enhance model generalizability and provide richer insights into regional and demographic variations. Real-time learning dashboards may be integrated to support continuous monitoring, while reinforcement-learning methods could optimize personalized study plans or remedial schedules. These expansions would further strengthen the system's ability to support large-scale educational decision-making.

References

[1] A. Al-Barrak, M. Al-Razgan, and N. Al-Shehri, "Student Academic Performance Prediction," *International Journal of Computer Applications*, vol. 25, pp. 21–28, 2010.

[2] K. VanLehn, "The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems," *Educational Psychologist*, vol. 46, no. 4, pp. 197–221, 2011.

[3] R. Jain, S. Salunke, and K. Mane, "Predicting Academic Performance Using Decision Trees," *IEEE Access*, vol. 3, pp. 1–8, 2015.

[4] D. K. Kolo, S. A. Adepoju, and J. Alhassan, "A Decision Tree Approach for Predicting Students' Academic Performance," *International Journal of Education and Management Engineering*, vol. 5, no. 5, pp. 1–9, 2015.

[5] V. Kumar and A. Singh, *Hybrid Machine Learning Models for Dropout Prediction*. Springer, 2018, pp. 89–99.

[6] W. Holmes, M. Bialik, and C. Fadel, *Artificial Intelligence in Education: Promises and Implications for Teaching and Learning*. Center for Curriculum Redesign, 2019.

[7] S. Chakraborty, P. Roy, and A. Das, "Socio-Economic Factors in Educational Data Mining," *International Journal of Educational Data Mining*, vol. 12, pp. 55–64, 2020.

[8] V. Matzavela, C. Karagiannidis, and S. Argyropoulos, "Decision Tree Learning Through a Predictive Model for Personalization of Student Academic Performance," *Education and Information Technologies*, vol. 26, pp. 4821–4837, 2021.

[9] UNESCO, *Artificial Intelligence and Education: Guidance for Policy-Makers*. United Nations Educational, Scientific and Cultural Organization, 2021.

[10] P. Anderson, *Outcome-Based Education and Predictive Analytics*. Springer, 2022.

[11] J. Smith, R. Lee, and M. Khan, "Deep Learning for Employability Analytics," *IEEE Transactions on Learning Technologies*, vol. 16, pp. 445–456, 2023.

[12] S. Wang *et al*., "Artificial Intelligence in Education: A Systematic Literature Review," *Expert Systems with Applications*, vol. 244, Art. no. 122823, 2024.

[13] UNESCO, *AI in Education Report*. UNESCO Publishing, 2024.

[14] M. Shoaib *et al*., "AI Student Success Predictor: Enhancing Personalized Learning Using Advanced Machine Learning Algorithms," *Computers in Human Behavior*, 2024. [Online]. Available: ScienceDirect.

[15] M. D. Adewale *et al*., "Impact of Artificial Intelligence Adoption on Students," *Education and Information Technologies*, 2024. [Online]. Available: PMC.

[16] E. Ahmed *et al*., "Student Performance Prediction Using Machine Learning," *Computational Intelligence and Neuroscience*, 2024.

[17] V. K. Vijayakumar and A. Naveen, "Long-Term Effects of AI-Personalized Learning on Engagement and Performance," *International Journal of Environmental Sciences*, pp. 1601–1608, 2025.

[18] W. Ahmed *et al*., "Machine Learning-Based Academic Performance Prediction," *Scientific Reports*, Nature, 2025.

[19] W. Zou *et al*., "Prediction of Student Academic Performance Utilizing a Practical Framework with Multiple Machine Learning Models," *Applied Sciences*, vol. 15, no. 7, p. 3550, 2025.

[20] C. Merino-Campos, "The Impact of Artificial Intelligence on Personalized Learning in Higher Education: A Systematic Review," *Trends in Higher Education*, vol. 4, no. 2, p. 17, 2025.

[21] S. Hashemifar *et al*., "Personalized Student Knowledge Modeling for Future Material Prediction (KMaP)," arXiv preprint, arXiv:2505.14072, 2025.

[22] D. Brown, L. Carter, and M. Singh, "Data-Driven Continuous Improvement in Education," *Educational Data Mining Journal*, vol. 8, pp. 120–135, 2025.

[23] V. K. Vijayakumar and A. Naveen, "A Data-Driven Framework for Holistic Student Performance Evaluation and Industry Readiness Using Machine Learning," *Iconic Research and Engineering Journals*, vol. 9, no. 3, pp. 1350–1354, 2025.

[24] V. Krishnan, "An Analysis of the Impact of Artificial Intelligence Tools on Education," *Emperor Journal of Applied Scientific Research*, vol. 7, no. 2, pp. 1–14, 2025.

[25] V. Krishnan, "Current Practices, Benefits, and Challenges of Transforming Outcome-Based Education with Artificial Intelligence," in *Proceedings of the International Conference on Information, System and Convergence Applications*, Chennai, India, Aug. 9–11, 2024, ISBN: 978-81-972363-3-4.

[26] V. Krishnan, "AI-Driven Student Performance Prediction Using Multi-Class Classifiers," *International Journal of Innovative Science and Research Technology*, vol. 10, no. 9, pp. 2639–2645, 2025.

About the Author

Mr. Vijayakumar Krishnan is a research scholar at Don Bosco College (Co-Ed), Yelagiri Hills – 635 853, affiliated with Thiruvalluvar University, Vellore. He possesses over 9 years of experience in computer networking, server and IT administration, complemented by 3 years in academia and 1 year of research experience.

Dr. Naveen A serves as a Research Supervisor at Don Bosco College (Co-Ed), Yelagiri Hills – 635 853. With 7 years of combined academic and research experience, he currently mentors four research scholars under his guidance.

Acknowledgement

I express my sincere gratitude to all individuals and institutions who contributed to the successful completion of this research work. I extend my heartfelt thanks to my academic mentors and colleagues for their valuable guidance, constructive feedback, and continuous encouragement throughout the study.

I am also thankful to the participating institutions and the undergraduate students whose data and cooperation made this research possible. Their contributions were instrumental in developing and validating the proposed analytical framework.

My appreciation further goes to the technical and administrative staff who provided essential support during data collection, preprocessing, and analysis.

Finally, I convey my deep thanks to my family and well-wishers for their unwavering support and motivation, which enabled me to carry out this research with dedication and clarity of purpose.

Comment to the Editor

I respectfully submit this manuscript for your kind consideration. The research presented in this paper offers meaningful contributions to the field of educational analytics and student performance prediction. I believe the manuscript aligns well with the scope of your esteemed journal, and I sincerely appreciate the time and effort invested by the editorial team in reviewing this submission. I look forward to your feedback and am willing to make any revisions necessary to enhance the quality of the publication.